(12) UK Patent Application (19) GB (11) 2 283 645 (13) A

(43) Date of A Publication 10.05.1995

(21) Application No 9322910.2

(22) Date of Filing 06.11.1993

(51) INT CL[6]
H04L 12/46 12/66

(52) UK CL (Edition N)
H4P PPA

(56) Documents Cited
GB 2267418 A

(58) Field of Search
UK CL (Edition M) H4P PPA PPG
INT CL[5] H04L 12/46 12/66
Online databases:WPI,INSPEC

(71) Applicant(s)
Digital Equipment International Limited

(Incorporated in Switzerland)

1 Grand Places, 1700 Fribourg, Switzerland

(72) Inventor(s)
Stewart F Bryant
Ian Michael Charles Shand

(74) Agent and/or Address for Service
Eric Potter Clarkson
St Mary's Court, St Mary's Gate, NOTTINGHAM, NG1 1LE, United Kingdom

(54) Digital communication systems

(57) A digital communication system comprising a network of routers R1-R3 linked together by links LK12-LK23 and having LANs LAN1-LAN7 coupled to them, and using IP (Internet Protocol), under which each LAN has a subnet address, and each host on a LAN has the subnet address as the high-order part of its own address. In IP, each router contains a set of interface/LAN tables each listing the low-order address portions of the addresses of the hosts attached to the LAN plus the MAC (medium access control) identifiers of those hosts, and a set of link tables listing the subnet addresses of the LANs reachable through those links. In the present system, both the interface tables and the link tables contain the full host addresses of all hosts reachable through those interfaces and links, and the routers also contain means for polling the interfaces for unknown hosts. Each router also contain an ARP (address resolution protocol) unit (30, Fig. 2) for detecting ARP requests from a source for a destination having the same subnet address as the source but not on the same interface, and returning a proxy ARP response giving the router's identification. A host can thereby be moved to a LAN whose address does not match that of the host.

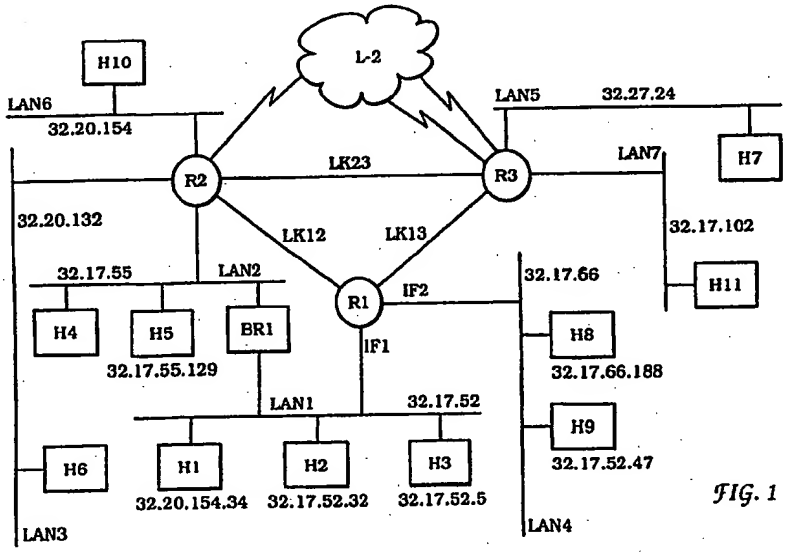

FIG. 1

GB 2 283 645 A

At least one drawing originally filed was informal and the print reproduced here is taken from a later filed formal copy.

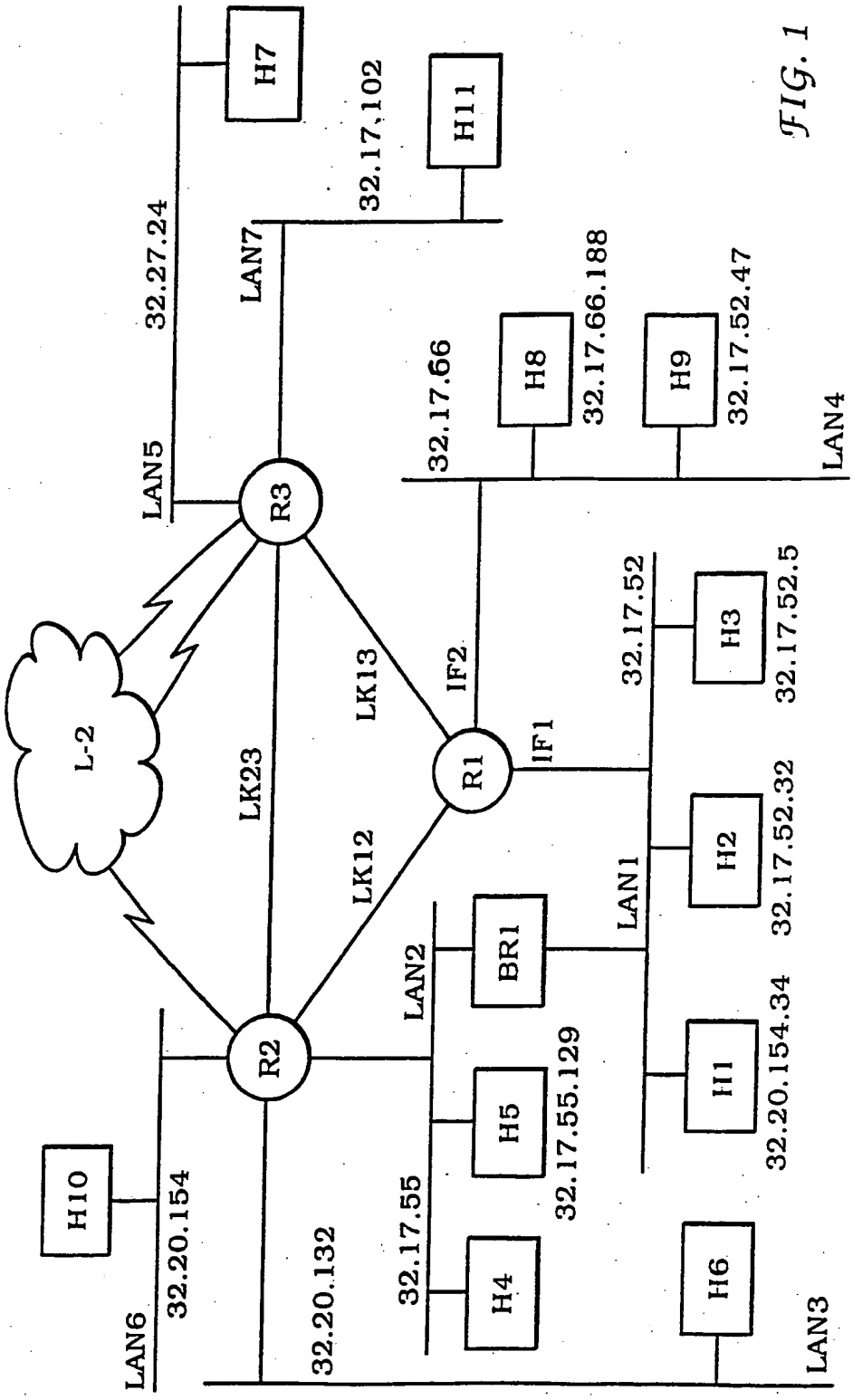
H7
H11
32.17.102
32.27.24
LAN7
LAN5
R3
32.17.66
H8
32.17.66.188
H9
32.17.52.47
LAN4
L-2
LK23
LK13
IF2
R1
IF1
32.17.52
H3
32.17.52.5
LK12
LAN1
H2
32.17.52.32
LAN2
BR1
H10
R2
H1
32.20.154.34
32.17.55.129
H5
32.17.55
H4
32.20.154
32.20.132
H6
LAN6
LAN3

FIG. 1

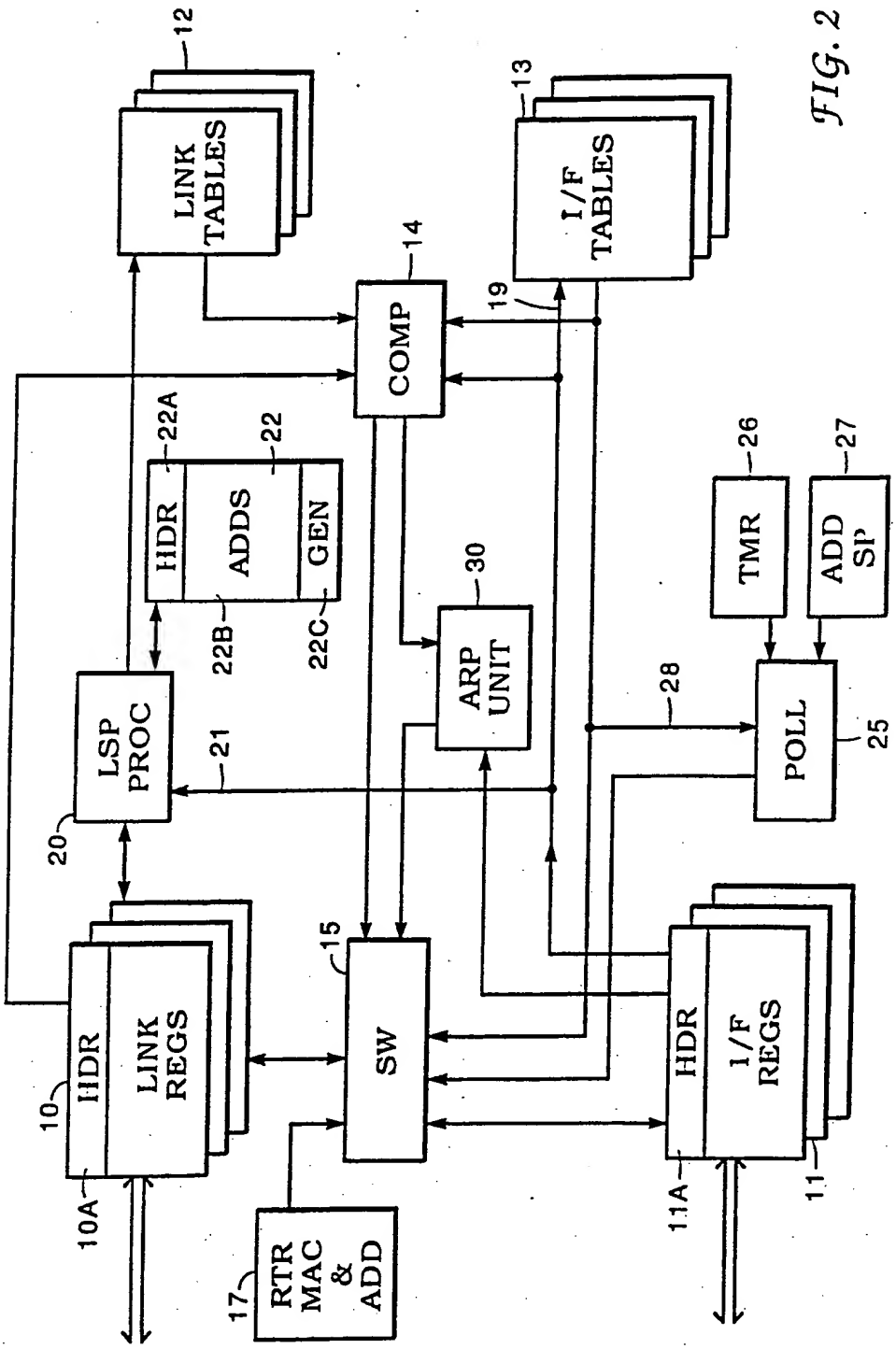
12
LINK TABLES
13
I/F TABLES
14
COMP
19
22A
22
HDR
ADDS
GEN
22B
22C
30
ARP UNIT
26
TMR
27
ADD SP
28
LSP PROC
21
20
POLL
25
15
SW
10
HDR
LINK REGS
10A
HDR
I/F REGS
11A
11
RTR MAC & ADD
17

FIG. 2

Digital Communication Systems

The present invention relates to digital communication systems, and more particularly to the addressing of units therein.

Digital communication systems: general

There is a considerable variety of digital communication systems. We are primarily concerned here with systems which interconnect a considerable number of essentially independent units (typically devices such as personal computers and work-stations), and are typically geographically extensive. Depending on the particular type of system, the units which it interconnects are termed end units, end-systems, or hosts.

As a very general matter, there are two extreme forms of system: a pure switching system and a pure broadcast-type system. In a pure switching system, the connections between the hosts are all individual, passing through a network of switching nodes. In a pure broadcast-type system, all end-units are connected to all other end-units by means of a common message medium.

It is clear that both these extreme types of system have major disadvantages. A pure switching system requires a highly complicated network of switching nodes, while there are obvious capacity limits on a pure broadcast-type system. A hybrid style of system has therefore become well established, in which there are local broadcast-type subsystems which are connected to each other by means of a switching system. (In a sense, this constitutes a hierarchy, but the term "hierarchy", and the associated term "levels", are normally used to describe the organization of the more complicated and elaborate forms of switching network.)

A simple and common form of local broadcast-type subsystem is the LAN (local area network). A LAN consists essentially of a common message medium to which a number of hosts are connected. When a host wants to send a message, it monitors the LAN to determine whether any other host is currently using the LAN. If not, then the host sends its message. Every host permanently monitors the LAN, watching to see whether any of the messages on the LAN are directed to itself. (There are various mechanisms for dealing with collisions, where two hosts try to transmit substantially simultaneously.)

There are various specific forms and various modifications of LANs, and there are other similar broadcast (common medium) systems. We shall use the term LAN loosely to cover all such systems, regardless of the details of the manufacturer or protocol.

As noted above, a number of LANs may be coupled together or interconnected by means of a switching network. The switching network in general consists of a number of nodes or switching devices, which we shall term "routers". (Alternative terms are "intermediate systems" and "gateways".) The connection from a router to a LAN is termed an interface; the connection from a router to another router is termed a link.

Obviously, there must be a suitable addressing system. Each host must have an address, and the communication system must somehow deliver messages from any host to any other host.

IP systems

We will consider primarily the types of system known as IP (Internet Protocol) systems from now on, because that is the main type of system for which the present invention is applicable. However, the principles of the

present invention are not limited to IP systems, but are also applicable to other systems having similar characteristics, such as Appletalk.

Identifiers and addresses

Hosts generally have unique identifiers which are physically defined by their manufacturers, eg. by hard-wiring or burning in, often termed MAC (medium access control) identifiers. A MAC identifier is normally globally unique; it will typically include a portion distinguishing the manufacturer from all other manufacturers and a serial number distinguishing it from all other machines made by that manufacturer.

It is however preferred to assign each machine a logical address, which can be chosen to facilitate the finding of connection paths in the system. (If desired, a single physical machine can be given more than one logical address, in which case it will behave as more than one logical host.) The MAC identifier is more usually termed a MAC address, but we will use the term "identifier" for MAC addresses to avoid confusion with logical addresses.

In the IP system, the logical address (IP address) is a 32-bit number, which is conventionally divided into 4 bytes or octets which are then written in decimal form (eg. 1.5.21.178). These logical addresses are normally assigned manually.

A major feature of the IP system is that all hosts attached to a LAN have a common high-order part (which is typically the top 3 bytes) of their addresses; this high-order part thus forms the address of the LAN. Thus the host with address 1.5.21.178 will be attached to a LAN with address 1.5.21, and all the other devices attached to that LAN will have addresses with the same high-order part, eg. 1.5.21.17, 1.5.21.8, etc. Each LAN forms a subnet; the address of the LAN is normally termed a subnet address.

If only part of an address is significant, then the significant part is indicated by a mask associated with the address (and of the same length as the address). Thus for the above LAN, the mask will be 255.255.255.0, because only the top 3 bytes of the address are significant. (In theory, the mask can be used to define non-contiguous bits, but in practice this rarely happens.)

The IP system as so far described therefore consists of a network of routers with LANs attached to the routers. A host on a LAN can send messages to other hosts on the same LAN directly over the LAN. To send a message to a host on a different LAN, the host must send the message to the router attached to the LAN. The network of routers then has the responsibility for passing the message to the router attached to the destination LAN. That router then puts the message on that LAN, and the destination host receives it.

System elaborations

There are certain elaborations of this basic system which are worth noting.

First, there is a special address used for broadcasting. In effect, every device has two addresses; its own normal address and the special broadcast address. A message with the broadcast address is received and accepted by every device. Messages with the broadcast address are normally confined to a single LAN; the routers do not attempt to pass such messages through the router network. (There are in fact various special addresses, to allow multicasting (to a group, but not all, of the hosts), and a second broadcast address, but this is not relevant for present purposes.)

Second, a LAN can be connected to more than one router. This may be the most convenient way to connect two parts of the system, with the LAN forming the only connection between the two routers. More often, however,

the two routers are both in the same router network, so that they provide two alternative paths to the LAN (from some other LAN). This redundancy allows the system to maintain communication with the LAN even if one of the routers connected to the LAN fails; also, it may allow the message flow rate to or from the LAN to be increased above the limit attainable with one router.

Third, two LANs can be coupled together by means of a bridge in known manner. A bridge is, in effect, a relay device which repeats any message on either of the LANs onto the other LAN. Thus LANs can be connected together into an extended LAN network in manner well known in the art. (We shall use the simple term "LAN" to include extended LANs.)

Fourth, a LAN (which may be a single or extended LAN) can have more than one logical address; the router to which it is attached will treat the single physical port or interface to which that LAN is attached as two separate logical interfaces. Any message put onto such a LAN at any point is physically transmitted to all hosts on it. (In fact, a bridge may have some form of filtering built into it, but this is not relevant for present purposes.) Logically, however, the LAN consists of two or more distinct subnets with different subnet addresses.

IP message flow protocol on LANs

In the IP system, message routing through the router network is determined by the IP addresses, but message routing over LANs is determined by the MAC identifiers. Mechanisms are therefore required to convert IP addresses to MAC identifiers when messages pass over LANs.

There are three main cases to consider for a source host sending a message to a destination host: the destination host may be on the same subnet

as the source host, it may be on a different LAN, or it may be on the same LAN but a different subnet.

In the first case, the source sends out an ARP (address resolution protocol) request message with the logical (IP) address of the destination. That ARP request is received by the destination, which sends back an ARP response message to the source. (The ARP request is a broadcast message which is received by all the hosts on the subnet, but only the destination host responds; all the other hosts recognize that the destination address in the ARP request does not match their own address, and they therefore discard the message.) The destination host includes its MAC identifier in its ARP response. The source then sends the actual data message to the destination using the destination's MAC identifier.

This involves a large message overhead, since the passing of each data message is preceded by an ARP request and ARP response. The various units of the system therefore store tables of IP (logical) addresses and MAC identifiers, so that most data messages can be sent out with the MAC identifiers without having to be preceded by ARP requests and responses.

In the second case, the message has to be passed through the router network. In general, each host knows of the existence of at least one router on its LAN. (This may be achieved by routers advertising their presence to their hosts by means of broadcast messages.) The source therefore sends the message to a router (using the router's MAC identifier). The router network forwards the message to a router attached to the LAN including the destination host. That router then sends an ARP request to the destination host, which returns an ARP response. The router then sends the data message to the destination host, using the host's MAC identifier. (If the router does not get

an ARP response, the packet is discarded or sent back to the source with an error status.)

The third case is where the source and destination hosts are on the same LAN but different subnets. The first message is passed in the same way as for the second case; the router accepts the message and then transmits it again on the same LAN. The router also returns a redirect message to the source, informing the source of the MAC address of the destination. The source stores this information, and can then send any further messages direct to the destination over the LAN common to the source and destination.

More specifically, each host maintains a connection table which lists the IP addresses and corresponding MAC identifiers of other hosts with which it has recently been in communication (without passing through a router - ie. in the first and third cases above). If a host wants to send a message, the destination is initially identified by its IP address. The host checks its table for the IP address, and if it is in the table, it extracts the associated MAC identifier from the table and sends the data message directly to that MAC identifier. If the IP address is not in the table, then the host has to send an ARP request to obtain a MAC identifier for the data message to be sent to. It enters the MAC identifier and associated IP address in the table for future use.

Similarly, each router has a set of interface tables, one for each interface. Each table lists the logical subnet addresses for that interface and, for each subnet address, lists the hosts with that subnet address, by logical address and MAC identifier. Obviously, the router will only know of the hosts which have sent out ARP requests. Each table in the router also has the identifier of its physical interface associated with it.

The various tables normally incorporates a time-out mechanism, so that entries which have not been used for some considerable time are deleted. This minimizes the chance of a unit trying to send a message to a unit which has disappeared from the system.

Router network organization

As noted above, if a message has to pass through the router network, that network has the responsibility for passing the message to the router attached to the destination LAN. This means that the routers have to pass routing (addressing) information between themselves so that when a router on a LAN receives a message (from a host on that LAN) for another LAN, it will know how to forward the message through the router network (and similarly, the router to which the message is forwarded will in turn know which router to forward the message to, and so on throughout the router network). This routing information is passed between the routers by means of routing control messages.

We are assuming here that a subnet address consists of the top 3 bytes of a 32-bit IP address (as determined by the associated mask), and that all hosts on that subnet have the subnet address as the high-order part of their own addresses. (The routing control messages will also generally contain other information, eg. about the cost and capacity of the paths between routers.) The routers therefore only have to deal with subnet addresses.

We are primarily concerned with the type of network in which every router is in communication, directly or indirectly, with all other routers on essentially the same basis. This is a single-level (level 1) system, and the number of routers will generally be fairly modest for such a system. Various types of routing control mechanisms are known for achieving this; for convenience, we shall assume that the routing control messages are link state

packets (LSPs), and that the router network passes the LSPs around so that each router maintains a set of link tables, one for each link to other routers, with each subnet address being held in the table for a link which points to a router which is in some sense nearer to the actual location of that subnet.

Further details of the mechanisms which the router network uses to pass messages (either data messages passing between hosts, or router network control messages) through itself are not relevant for present purposes.

Router network elaborations

This basic mechanism for establishing and maintaining the topology of the router network is subject to possible elaborations.

As noted above, some of the links in the router network may pass through LANs. This does not affect the operation of the system, though of course any messages in the router network which pass through such links have to be encapsulated by the LAN messaging mechanism for their passage over those links.

The routers can operate algorithms for combining (condensing) subnet addresses. Thus if a router has several LANs attached to it with the same top 2 bytes in their addresses (eg. subnet addresses 1.5.21, 1.5.34, 1.5.26, etc), and these LANs are the only ones in the system with these top 2 bytes, that router can identify all those LANs by the single address of just the top 2 bytes (1.5). (As noted above, a mask will define the address as consisting of only 2 bytes.)

This mechanism allows the router network to be hierarchical. In each local (level 1) region of the network, the routers will have full information about all the LANs attached to that region, but will have only summary

(condensed) information about the subnet addresses of other local regions. The mechanisms used for passing messages between different level 1 regions of the router network form a second level, level 2, of the network.

The condition for combining subnet addresses in a router in a level 1 area can in fact be relaxed slightly. If there is a odd subnet with say 1.5.102 as its address attached to some other router, the router to which all other subnets with 1.5 as the top 2 bytes in their addresses are connected can advertise itself as the router for address 1.5, provided that it forwards any messages it receives for subnet 1.5.102 on to the other router having the 1.5.102 subnet attached to it.

The problem

In the IP system, the address of a host includes the address of its LAN (as a subnet address). A host is therefore "tied" to its LAN. It can be moved to a physically different place on its LAN; all physical locations on a LAN are logically identical. However, it cannot be moved to another LAN. If it is so moved, it will be inaccessible; although it will be physically attached to the new LAN, no other host will be able to reach it.

In some situations, this restriction on moving hosts is not significant; in others, it provides a useful security feature.

However, in a large company or other organization, there may be a number of different LANs which are connected in an IP system, and for a variety of reasons, such as changes of organization, it may be desirable or even necessary to physically move a host in such a way that it has to be removed from its LAN and attached to another LAN.

This causes a problem. To move a host from one LAN to another, the host's address has to be changed to match the address of the new LAN. Since addresses are manually assigned, it is possible to make this change. Making the change may not in itself be particularly difficult. However, that will in effect turn the host into a new host. None of the other hosts which have been in communication with it will know its new address, and communication will have to be re-established from scratch with all these other hosts. This can be highly inconvenient.

One potential solution to this difficulty is to couple the different LANs together to form an extended LAN, as discussed above. However, this increases the complexity of system management, and involves difficulties arising from the complexity and proprietary nature of multipath bridging. The message density on the extended LAN is also increased, eg. by the increased multicast traffic, and this may limit the extent to which this solution is feasible.

Another potential solution utilizes a directory service. IP systems often have a directory service, which is essentially a table correlating host "names" with their IP addresses. This allows a source host to identify a destination host by means of the destination host's name; but before a source host can actually communicate with a destination host, the source has to obtain the destination's IP address from the directory service by sending the destination's name to the directory service, which returns the associated IP address.

If a host is moved, it can be given a new IP address consistent with its new location, and the directory service can be updated to associate the new IP address with the host's name (which is unchanged). If a source host wants to communicate with the host which has moved, the source host will find that messages directed to the old destination IP address will fail to reach their destination. It can then use the directory service to obtain the destination's IP

address, as if it were trying to establish communication with the destination for the first time, and will thereby acquire the destination's new IP address.

This solution requires manual updating of the directory service, which is likely to involve considerable time delays during which the migrated host is inaccessible. It also requires the organization to have a management structure capable of dealing with the changes involved in an acceptably simple and effective manner.

A third potential solution is to provide re-addressing. This involves giving the migrated host a new IP address, consistent with its new LAN, and recording its old and new addresses in the router for its original LAN. A message sent to the host using its old address will reach its old router; that router will replace the old address by the new address and forward the message to the new router. However, this has various disadvantages. For example, message paths through the router network are considerably extended; also, the number of host addresses used in the system is increased each time a host migrates, and the need for the migrating host to be given an address consistent with its new LAN may be inconvenient. Also, the return path for messages between the two hosts is different to the outward path, which can cause difficulties.

The broad object of the present invention is therefore to provide an improved technique whereby a host in an IP or similar system can be moved from one subnet to another without having to have its address changed.

There are some important constraints implied in this formulation of the problem. Any solution must be compatible with existing IP systems; any modifications to only some of the routers and/or hosts in an existing IP system to provide the required technique must not interfere with the operation of the

remaining routers and/or hosts. Further, IP-type systems are of course very well established, and include huge numbers of existing hosts. It is therefore desirable, if possible, for the solution to involve modifications to only routers, so that existing hosts can be moved without having to be modified.

As just noted, IP-type systems are very well established, and many such systems are extremely large, both in the numbers of hosts and geographically. The ideal solution in an abstract sense would permit a host to be moved from any location on the system to any other location. However, a solution which allowed only a limited degree of mobility of hosts around the system would be of great practical value, even though it would theoretically be only a partial solution.

The solution

The present invention provides a solution which comprises a combination of several features, all involving modifications of the details of the manner in which the routers operate.

According to the present invention there is provided a digital communication system comprising a network of routers linked together by links and having interfaces with local area networks (LANs) coupled to them, and operating under a protocol under which each LAN has a subnet address, and each host on a LAN has the subnet address as the high-order part of its own address, each router containing a set of interface/LAN tables listing the low-order address portions of the addresses of the hosts attached to the LAN plus the MAC (medium access control) identifiers of those hosts, and a set of link tables listing the subnet addresses of the LANs reachable through those links, wherein: both the interface tables and the link tables in the routers contain the full addresses of all hosts reachable through those interfaces and links; the routers contain means for detecting ARP (address resolution protocol) requests

from a source host for a destination host having the same subnet address as the source host but not on the same interface, and returning a proxy ARP response giving the router's identification; and the routers contain polling means for polling the interfaces for unknown hosts.

Since the present system does not require any changes to the hosts, the address space of the system is unchanged. However, the present system in effect decouples the host addresses from the subnet addresses and hence from the geographical LAN locations. This allows considerably greater freedom in assigning addresses to hosts within the system address space.

In the standard system as described above, we have taken the top 3 bytes of the 32-bit address space as being used for subnet addresses, and the bottom byte as being used for different host addresses on the subnet. In fact, the division between the subnet address and the host addresses on the subnet can be defined more flexibly, by the use of suitable masks. However, the number of possible host addresses on a subnet must obviously be a power of 2, and the actual number of hosts on the subnet is likely to fall well short of the maximum.

Thus in the standard system, there are likely to be many spare addresses, which cannot be used (or can only be used by hosts added to the subnet with which those unused addresses are associated). In the present system, these spare addresses can be used much more freely, since they can be assigned to hosts regardless of which subnets (and hence LANs) those hosts are to be attached to.

Router network organization

A major feature of router operation is that the present modified routers use full host addresses for level 1 routing.

In the standard IP system, the routers use abbreviated addresses - the subnet addresses of the LANs - for level 1 routing; in effect, the routers operate an address compression algorithm which compresses the addresses of all the hosts on a subnet into a single subnet address. In the present system, the router operation is modified so that this address compression is no longer performed.

The result is that in a router network using the present modified routers, each router will hold effectively the same routing information as before, albeit in an expanded form. The operation of the router network is therefore effectively unchanged in principle (as far as the routing of messages through the router network is concerned). However, the amount of LSP traffic is increased, the amount of processing required for routing is increased, and the routers have to have a greater storage capacity.

If the level 1 network forms part of a larger system coupled to other level 1 networks through a level 2 organization, the level 2 organization is unaffected. Compressed or summary addresses are used unchanged for level 2 routing. The migration of hosts is restricted to within their own level 1 systems; it is not possible for a host to migrate from one level 1 system to another. As mentioned above, this restriction is rarely significant.

The present routers are largely compatible with standard routers, so that a network can consist of a mixture of standard and compatible routers. For present purposes, it is convenient to regard the resulting network as a network of modified routers to which standard routers have been added.

In a standard router network, the routing information consists of subnet addresses, which are distributed by the LSPs and stored by the routers. A subnet address has the form of a full address plus a mask, with the mask

defining which part of the full address forms the subnet address; the rest of the address is ignored. The ignored part of the address is in fact, of course, a host address (on the subnet defined by the mask).

In a mixed system, the modified routers will send LSPs with full addresses in the same format, ie. address plus mask pairs, and any standard router receiving such an LSP will automatically store this address in the usual way. As far as such a standard router is concerned, there is no difference between subnet and full (host) addresses; the distinction arises solely from the contents of the masks associated with the various addresses. In a mixed system, therefore, the presence of standard routers will not affect the performance of the subsystem of modified routers (provided, of course, that the standard routers have sufficient storage capacity). The migration of hosts in such a mixed system is of course limited to the subsystem of modified routers.

As noted above, in the present system the amount of address information which has to be propagated through the router system is considerably increased. It may therefore be desirable to introduce a new LSP option type or format, to reduce the size and/or number of LSPs.

In the standard system, routing information is exchanged between the routers in the form of information units termed "options", of which there can be various formats or types. To reduce the number of LSP messages, a number of options are typically assembled into a single LSP. The standard option type can be taken as consisting of a header, an address section, and a general information section. The header contains an identifier which defines the LSP option type and length; the address section will consist of the address plus mask pair; and the general information section will contain associated routing information such as cost and distance.

The new LSP option type has the same general format, but the address section contains a set of host addresses without masks. Thus a considerable number of addresses can be sent as a single option of the new type, instead of needing a separate option (of the old type) in the LSP for each address. The length of this new LSP option will be considerably less than the total length of the separate LSP options of the old type, because there will be only one header and general information section, and each address will consist of a pure address with no accompanying mask.

The number of addresses in the new LSP option type may be included explicitly in the header, or may be calculated from the total length of the option by subtracting the header and general information lengths and dividing by the address length. Different addresses cannot, of course, have different associated routing information, because the addresses all share the common general information in the final section of the option. The routers will normally assemble the addresses of hosts on a common LAN when constructing an LSP option of the new type; those addresses will then all have the same characteristics and can share the same general information.

The new LSP option type can be used in a mixed system, as standard routers forward all LSP options (including those of the new type); the full host address information will thus be maintained throughout the subsystem of modified routers. However, the standard routers will not update themselves with the contents of LSP options of the new type. The modified routers must therefore also send out LSP options of the old type, so that the routing information in the standard routers is maintained. Also, if the standard routers split the subsystem of modified routers into disconnected parts, hosts cannot migrate between those parts because the standard routers connecting those parts will maintain only subnet addresses, not full host addresses.

LAN addressing by routers

For the router network to be able to route messages correctly, each router must know the whereabouts of all the hosts. This knowledge is distributed amongst the different routers using LSPs. However, before the a router can distribute information about the location of a host, it must become aware of the existence of the host.

In the standard system, each router is aware of the subnets attached to it (this knowledge may, for example, be entered manually). A router need not be explicitly aware of the existence of the hosts attached to the LAN. If a message for a host on the LAN is received, the router can send out an ARP request, and the ARP response confirms the existence of the host on the LAN. (If there is no ARP response, then it is assumed that the host does not exist.)

As discussed above, in the standard system the router in fact maintains an interface table for the various hosts on the LAN, so that it can forward future messages to them without having to obtain their MAC identifiers by ARP requests. The interface table is built up partly from ARP requests sent out by the router, and partly by ARP requests sent out from the hosts.

A standard router preferably maintains this table actively, by polling the hosts at suitable intervals. The poll message is simply an ARP request to the host. An ARP response confirms the existence of the host; if there is no ARP response, the host no longer exists.

In the present system, the modified router maintains its interface tables in broadly the same way as do standard routers. The present router, however, necessarily constructs its interface tables entirely automatically, whereas in a standard router the subnet addresses may be entered manually.

The present routers listen promiscuously for ARP requests from hosts (for a reason discussed later), and may listen similarly to other messages. This listening helps the routers to maintain their interface tables. The routers will therefore automatically learn of the existence of hosts which are involved in message transmission. However, it is possible that when a host is moved from one router to another, some other host may want to send a message to it before it has itself tried to send any messages. (In the standard IP system, hosts do not advertise themselves; the identity between the subnet address and the top part of the host address means that the location of a host is inherent in the system.)

Another way in which the routers can automatically discover the existence of hosts is for the hosts to announce their existence when they are first turned on, with the routers listening for such messages. However, this requires the hosts to issue suitable identification messages when first turned on; this may require modification of some hosts, and some types of host may not be modifiable.

Some mechanism must therefore be provided for the router network to discover the existence of silent hosts. Since changes cannot be imposed on the hosts themselves, the standard routers must therefore be modified, in the present system, so that they can discover the existence of such silent hosts. There are two ways in which this can be done, which may be termed active and passive. (It may be noted that in standard routers, the subnet addresses may be passed round automatically but will normally be set manually. In the present system, the subnet addresses are not of such importance, and the modified routers must determine all host addresses automatically.)

With the active technique, the routers actively search for hosts. Each router has to be modified to perform polling. For this, the routers in the level

1 area are informed (eg. manually) of the address space of the hosts in that area. Each router then polls each of its interfaces in turn; that is, for each interface in turn it sends out a series of ARP requests, working through the host address space address by address. It will therefore elicit responses from all hosts attached to it.

This polling is of course distinct from the polling, mentioned above, which standard routers perform. The standard router polling is not through the host address space, but through the actual addresses of hosts which are already recorded in the routers' tables, to confirm their existence.

Provided that the host address space is manageably small, this is the preferred mechanism. This polling automatically takes care of the normal maintenance of the interface tables. Since the router network can only route messages to hosts which it knows about, it is important to confirm the disappearance of a host, eg. by a suitable number of retries. A discovery time of say 500 s (comparable to the ARP time-out), and a polling rate of 7 polls per second will accommodate an address space of 4000 hosts, which is much larger than most practical LANs.

The polling message density can be reduced if a router does not poll for addresses which it knows to be attached to other of its interfaces, or to other routers. However, a router needs to poll for hosts which are attached to it to confirm their existence, just as with a standard router; also, polling for hosts which are listed in its tables as being attached to other routers accelerates their discovery if they are moved.

Instead of polling by ARP requests, a router could poll by sending a suitable broadcast message, asking the hosts to report their existence. However, this has two disadvantages. One is that it requires the hosts to return

suitable identification messages in response to the broadcast enquiry; this may require modification of some hosts, and some types of host may not be modifiable. The other is that the response messages from the hosts will temporarily produce a very high message density, which may for example overwhelm the router.

With the preferred mechanism of ARP polling, the polling intensity can be reduced by partitioning the host address space so that certain segments of it will only contain hosts which will announce their presence when first turned on. It will then not be necessary to poll through those address space segments.

A possible refinement of ARP polling is that if a router discovers that a host has disappeared, that host address can be distributed to all routers, with all routers then sending out ARP requests at higher than normal polling frequency for that host for some convenient period of time. (The router which has lost the host should be included in this, because the host may be migrating to another of its interfaces.) This will result in rapid detection of the migrating host if it is reconnected into the network.

With the passive technique for routers to discover the existence of silent hosts, they only search for a host when there is a message for that host. If the router network receives a message for a host which it (the router network) does not recognize, then the message is passed around the routers, and each router polls each of its interfaces with an ARP request.

This requires a more complex router network organization, to ensure that the message is distributed to all routers, but it may reduce the amount of polling, as the occurrence of messages to silent hosts will usually be relatively uncommon. The message may be distributed rapidly to all routers, with all the routers then polling their LANs; this may impose a significant transient load on

the system. Alternatively, each router in turn may poll its interfaces for the destination host, and forward the message on to the next router only if it fails to find the destination host on any of its interfaces; this may result in a large delay.

Host to host communication

In the standard system, there are 4 mechanisms for a source host to send a message to a destination host. First, if the destination has the same subnet address as the source, if the source does not know the destination's MAC identifier it will send an ARP request to the destination; otherwise (second), it sends the message direct to the destination using the MAC identifier. Third, if the destination is on a different subnet, the source sends the message to a router. Fourth, if the destination is on a different subnet but the same extended LAN as the source, the source can send direct to the destination's MAC identifier as a result of a redirect message from a router.

The present system must maintain all these modes of message transmission as far as the hosts are concerned; in particular, it must cope with all possible combinations of source and destination subnet addresses and LAN locations. The source and destination may be on the same or different LANs, and may have the same or different subnet addresses.

If the source and destination are on the same LAN and have the same subnet address, then if the source knows the destination's MAC identifier it will send the message direct to the destination using the MAC identifier. Otherwise, the source will send an ARP request to the destination and, because the two are on the same LAN, it will get an ARP response and then send the message using the MAC identifier returned in the ARP response. This is the same as in the standard system.

If the source and destination are on different LANs and have different subnet addresses, the source will send the message to a router, which will forward it to the router to which the destination is attached. This is broadly similar to the standard system (though the router uses the more detailed routing information of the present system).

If the source and destination are on the same LAN but have different subnet addresses, the source will send the message to a router; this will return a redirect message to the source, which will then send the message direct to the destination using the destination's MAC address. This is broadly similar to the standard system (though again the router uses the more detailed routing information of the present system).

If the source and destination are on different LANs but have the same subnet address, then the source will send an ARP request to the destination, expecting to receive an ARP response with the destination's MAC identifier. The router on the source LAN must listen for such ARP requests and return ARP responses (this is the promiscuous listening for ARP requests mentioned above). On hearing an ARP request on an interface, the router must check its link tables and its interface tables for its other interfaces for the destination. If the destination is in those tables, it is in fact on a different LAN from the source. However, the source is expecting an ARP response. The router must therefore return a proxy ARP response - ie, it must return an ARP response on behalf of the destination. This proxy ARP response will of course contain the router's MAC identifier. The source will then send the message to the router, which must then forward it through the router network.

Specific Embodiment

A communication system embodying the invention will now be described, by way of example, with reference to the drawings, in which:

Fig. 1 is a general block diagram of the system; and

Fig. 2 is a highly simplified block diagram of a modified router (ie. the present router).

Fig. 1 shows a communication system with various typical features. The system consists of a level 1 network of 3 routers R1-R3 coupled by links LK12, LK23, and LK13 (the digits indicating the routers which each link couples together). This level 1 network forms part of a level 2 system (the rest of which is shown merely as a cloud L-2), and is coupled to the rest of the level 2 network by links shown as zig-zag lines.

Router R1 has 2 physical LAN interfaces, with LAN1 (with subnet address 32.17.52) and LAN4 (with subnet address 32.17.66) coupled to them; router R2 has 3 LAN interfaces, with LAN2 (with subnet address 32.17.55), LAN3 (with subnet address 32.20.132), and LAN6 (with subnet address 32.20.154) connected to them; and router R3 has 2 LAN interfaces, with LAN5 (with subnet address 32.27.24) and LAN7 (with subnet address 32.17.102) connected to them. LAN1 and LAN2 are connected together in known manner through a bridge BR1, forming a single extended LAN with two subnet addresses.

Hosts H1-H11 are coupled to the various LANs as shown. Each host has an address consisting of 4 bytes. In the standard system, each host's address will be the address of its LAN plus a final byte added to the end of the subnet address, as shown for hosts H2 (address 32.17.52.32), H3 (address 32.17.52.5), H5 (address 32.17.55.129), and H8 (address 32.17.66.188).

Each host maintains a connection table for its connections. Host H2, for example, will maintain the following table:

Host H2, connection table

Router connection:
R1-MAC

Host list:

32.17.52.5 (H3)	H3-MAC
32.17.55.129 (H5)	H5-MAC

...

This table has two parts, a router connection and a host list. The router connection part is a single entry, the MAC of router R1, which the host uses for sending messages to hosts which are not on its own extended LAN. The second part lists the hosts which H2 has recently sent messages to, together with the MAC identifiers which it uses to send messages to those hosts. Communication with H3 is direct, over LAN1, so messages to that host are sent to that host's MAC identifier. Communication with H8 is via the router network, so messages to that host are sent to router R1, using that router's MAC identifier. Communication with H5 is also direct; H2 has learnt H5's MAC identifier as the result of a redirect message from router R1 or R2 at some time in the past.

The host maintains this table as a cache with time-out, so that entries which have not been used for more than a certain time are deleted. New entries are added as communication with new hosts is desired, by using the ARP requests as discussed above.

Each router maintains interface and link tables. Router R1, for example, would maintain the following interface tables if it were a standard router.

Router R1 (standard form), interface tables

IF1 (interface)
    32.17.52 (subnet address)
        5 H3-MAC
        32 H2-MAC
        ...
    32.17.55 (subnet address)
        129 H5-MAC
        ...

IF2 (interface)
    32.17.66 (subnet address)
        188 H8-MAC
        ...

Each interface table is divided into a separate section for each logical subnet address of the (possibly extended) LAN attached to that interface. Each section records the subnet address and then lists the hosts on the LAN with that subnet address. Each host entry consists of the host's address and its MAC identifier. The host's address is recorded as only the final byte, since the first 3 bytes of the address are the address of its subnet. The first interface table has two sections because the two LANs LAN1 and LAN2, with different subnet addresses, are both connected to that physical interface (via the bridge BR1 in the case of LAN2).

The routers also maintain link tables for their links to other routers. In the standard system, each router passes the addresses of the LANs to which it is coupled to the other routers in the level 1 network, and those other routers hold that information in their link tables. Thus if router R1 were standard, it would maintain two link tables as follows.

Router R1 (standard form), link tables

LK12:
32.20.154 (LAN6)
32.20.132 (LAN3)
...

LK13:
32.27.24 (LAN5)
32.17.102 (LAN7)
...

The entries in each table are the addresses of the LANs which can be reached through the associated link. If link LK13 did not exist, then the link table LK12 would contain the addresses 32.27.24 and 32.17.102 (as well as 32.17.154 and 32.20.132), because the message route for those addresses would then be via router R2.

Each subnet address in the link tables can be regarded as a compressed version of the set of host addresses on that LAN. This can be represented more fully by writing the subnet addresses as 32.17.154.xx, etc, where the final byte is masked off by a mask.

For the coupling to the rest of the level 2 system, the routers provide further compressed addresses over the zig-zag links to region L-2. In this case, the level 2 addresses will be simply the single value 32.0001xxxx.xx.xx (where the second byte is written in binary). Similarly, the routers R2 and R3 will maintain level 2 connection tables (with further compressed entries) for addresses in the L-2 region. Also, router R1 will maintain these region L-2 addresses (preferably in the same compressed form) in its link tables LK12 and LK13, so that messages from hosts on its LANs to the L-2 region can be correctly routed. (This is why the tables LK12 and LK13 are shown as having further entries beyond the 2 shown explicitly for each.)

In the present system, the routers are modified from the standard form to maintain the level 1 connection information within the interface and link tables in uncompressed form. Thus the interface table for interface IF1 for router R1 will contain:

Router R1 (modified form), interface table IF1

IF1:

32.17.52.5	H3-MAC
32.17.52.32	H2-MAC
32.17.55.129	H5-MAC
32.20.154.34	H1-MAC
...	

Switching, for convenience, to router R3 to discuss the link tables, this would contain the following link table for link LK13 in the standard form:

Router R3 (standard form), link table LK13

LK13:

32.17.52	(LAN1)
32.17.55	(LAN2)
32.17.66	(LAN4)
...	

For the modified form of router R3, this link table will contain all the entries for router R1's interface tables, in the same form as in those interface tables, instead of just the compressed or subnet addresses. Thus router R3 will contain the following link table for link LK13 in the modified form:

Router R3 (modified form), link table LK13

LK13:

32.17.52.5	H3-MAC
32.17.52.32	H2-MAC
32.17.55.129	H5-MAC
32.17.66.188	H8-MAC
...	

The connection information relating to connections to the level 2 area L-2 is unchanged from the standard form.

With the host addresses discussed up to now, the operation of the system with the present (modified) routers is substantially unchanged from the operation with standard routers. Suppose, however, that host H1 was originally on LAN6 (address 32.20.154), and was given the address 32.20.154.34 while it was on that LAN. Suppose also that it is desirable to transfer that host to LAN1 as shown.

In the standard system, it would not be possible for any messages to reach H1, because its address does not match LAN1's address. For H1 to be logically connected to the system, either its address would have to be changed to match that of LAN1 (so that it would effectively be a new host), or LAN1 and LAN6 would have to be coupled together by a bridge (so that H1 could still be reached by router R2), or some form of address conversion would have to be provided.

In the present system, however, messages can reach H1. This is because all the routers' tables (ie. both link and interface tables) contain the individual addresses of all hosts (of the level 1 area) in full; they do not now contain the subnet addresses as such.

Thus router R1 will contain the address 32.20.154.34 in its interface table for interface IF1, so that it can forward a message for H1 reaching it from another router (or from another of its interfaces). Similarly, the link tables of R2 and R3 will contain the address 32.20.154.34 in full, so that any message for H1 reaching either of those routers can be forwarded to router R1 (assuming that for some reason, messages are not passed to it from router R2 through the extended LAN network of the 2 LANs LAN2 and LAN1).

In the standard system, router R2 would contain the address 32.20.154 of LAN6 in one of its interface tables, and would capture all messages to any host with that as the first 3 bytes of its address. In the present system, however, router R2 contains only the addresses of the individual hosts attached to it, not the subnet address as such. It will therefore not capture any messages to host H1, ie. to address 32.20.154.34, and will therefore not interfere with the correct routing of messages to that host.

Fig. 1 also shows a second host, host H9, which has migrated, in this case from LAN1 to LAN4. Router R1's interface table for interface IF2 contains the address (32.17.52.47) of this host, and routers R2 and R3 contain this address in their link tables LK12 and LK13, so that messages to this host from LANs attached to R2 and R3 will reach it as desired.

There is however a complication if a host such as H2 on H9's original or "home" subnet wants to send a message to it. H2 finds that its own subnet address is the same as H9's subnet address, and therefore sends an ARP request to H9 on LAN1 (hosts' behaviour is unchanged from in the standard system).

As discussed above, the present routers listen to all ARP requests from hosts on their interfaces, to detect ARP requests for migrated hosts. When a router detects an ARP request, it checks the address of the destination host

against the contents of its tables (both the interface tables and the link tables). If the destination host is on the same interface as the source host, the router ignores the ARP request (the destination host will respond to the ARP request with an ARP response and message transmission will proceed normally). But if the destination is not on the same interface as the source, the router responds with a proxy ARP response which includes its own MAC identifier. The source will then send the message to the router, and the router then forwards the message to the actual location of the destination.

Thus router R1 will detect the ARP request from host H2 for host H9, find that host H9 is not on interface IF1, and return a proxy ARP response to H2. H2 will then send the message to the router, which will pass it to interface IF2 so that it reaches H9.

If host H2 has previously been in communication with host H9 over their original common LAN, H2 will of course still have H9's MAC address in its connection table, and will continue to use that MAC address when trying to send messages to H9; and when H9 migrates, H2 will find that H9 has apparently disappeared. H2 will thereupon flush its connection table (or at least the entry for H9), and attempt to re-open communication with H9 by sending an ARP request. Router R1 will return a proxy ARP response to this, as just discussed, and communication with H9 will therefore be re-established.

Fig. 2 shows the general logical organization of the preferred form of modified (present) router, which we may take as router R1.

There is a plurality of link registers 10, one per link, for receiving messages (including LSPs) coming in over links from other routers and for holding messages to be transmitted over those links. There is a plurality of interface registers 11, one per interface, for receiving messages coming in over

the router's interfaces and for holding messages to be transmitted over those interfaces. There is a plurality of link table stores 12, one per link, for storing the link tables discussed above. There is a plurality of interface table stores 13, one per interface, for storing the interface tables discussed above. The link and interface registers 10 and 11 are coupled to switching circuitry 15.

Each of the link registers 10 and interface registers 11 has a header section 10A, 11A respectively for containing header information including, for example, the source and destination addresses and (in the case of messages in the interface registers) MAC identifiers. When a message is received in one of these registers, its destination address is compared by a comparator 14 with the addresses in the link and interface tables and moved from its initial register to the appropriate register for output, ie. from a link register to another link register, from a link register to an interface register, or from an interface register to an interface register. In addition, if the message is moved into an interface register, the MAC identifier of the destination is copied over line 16 from the interface table into the header section of the interface register.

(In practice, the messages may be stored in a common memory, with pointers being used to identify different memory areas as the different registers, and the movement of a message from one register to another being achieved by changing the pointers. Also, the headers may be processed separately from the bodies of the messages.)

The interface registers 11 are also coupled to an ARP unit 30. All ARP requests on the LANs attached to the interfaces are received by the router, ie. are written into the interface registers 11. When an ARP request is so received, comparator 14 compares the host destination address in its header with the host addresses in the link tables 12 and the interface tables 13.

If the destination is not in the interface table for the interface on which the ARP request was received, ie. is in some other interface table, or in a link table, then the comparator 14 sends a signal to the ARP unit 30, which then converts the ARP request in the interface register to a proxy ARP response. This ARP response includes the router's address and MAC identifier, which are stored in a router address and MAC identifier store 17 and are copied into the header into of the interface register for return to the host as the ARP response.

If the destination is in the interface table for the interface on which the ARP request was received, then the router makes no response. However, for all ARP requests which it receives, the router checks whether the source is listed in the interface table for the interface on which the ARP request was received. If it is not, then it updates its tables by adding the source's address to the appropriate interface table and deleting it from any other tables which it is in. In addition, the host's address (ie. the full address) is passed (over line 21) to an LSP processor 20.

A polling unit 25 is also coupled, through the switching circuitry 15, to the header sections 11A of the interface registers 11. The polling unit 25 performs two functions, under the control of a timer 26 to which it is coupled.

First, the polling unit is coupled to the interface tables 13, and selects each entry in the interface tables in turn for verification. For this, the address of each end-station in turn is copied into the appropriate one of the interface registers 11 and sent out as an ARP request. The MAC identifier in the response is passed back to the interface table and compared therein with the stored MAC identifier, to verify the entry. If verification fails (after a suitable number of retries), the table entry is deleted and the address of the deleted host is passed to the LSP processor 20.

Second, the polling unit is also coupled to an address space store 27, which is set to contain the address space of the (level 1) system. Under control of the timer 26, the polling unit 25 works sequentially through all addresses of the address space. Addresses which are already in the interface tables are filtered out. The remaining addresses are passed, in sequence, to each of the interface registers for sending out an ARP request, to see whether a host with that address exists. If it does, then the address and MAC identifier in the ARP response are passed to the appropriate interface table and to the LSP processor 20.

Turning now to the LSP processor 20, this receives the addresses of hosts newly discovered by the router and of hosts which disappear from the router's interfaces. It constructs LSP options containing these addresses, assembles them into LSPs, and passes them to the set of link registers 10 for transmission to other routers. This processor 20 also processes LSP options received by the link registers 10 from other routers, updating the entries in the corresponding link table 12 by adding and/or deleting entries appropriately. The LSP processor 20 is coupled to an LSP option memory 22 in which LSP options of the new type discussed above are constructed; this memory comprises a header section 22A, an address section 22B for the addresses of the LSP, and a general information section 22C. This memory is used to assemble LSP options of the new type which are to be sent out by the router, and to store incoming new type options received from other routers ready for analysis and transfer of their contents into the link tables 12.

In the system shown in Fig. 1, each router is coupled to every other router. In general, however, this will not always be so. LSP options must therefore be forwarded throughout the level 1 area. The LSP processor is responsible for this; it causes an incoming LSP option to be copied to all other link registers 10 for forwarding (as parts of LSPs) to other routers. Various

techniques can be used to prevent the unlimited circulation and multiplication of LSP information.

Claims

1. A digital communication system comprising a network of routers linked together by links and having interfaces with local area networks (LANs) coupled to them, and operating under a protocol under which each LAN has a subnet address, and each host on a LAN has the subnet address as the high-order part of its own address, each router containing a set of interface/LAN tables listing the low-order address portions of the addresses of the hosts attached to the LAN plus the MAC (medium access control) identifiers of those hosts, and a set of link tables listing the subnet addresses of the LANs reachable through those links, wherein:

both the interface tables and the link tables in the routers contain the full addresses of all hosts reachable through those interfaces and links;

the routers contain means for detecting ARP (address resolution protocol) requests from a source host for a destination host having the same subnet address as the source host but not on the same interface, and returning a proxy ARP response giving the router's identification; and

the routers contain means for interrogating the interfaces for unknown hosts.

2. A digital communication system according to claim 1 wherein the means for interrogating the interfaces comprises polling means.

3. A digital communication system according to claim 2 wherein the polling means include timing means causing the polling means to perform polling for unknown hosts.

4. A digital communication system according to claim 3 wherein each router contains an address space store settable to contain the address space of the system.

5. A digital communication system according to claim 2 wherein the polling means of a router poll for unknown hosts is in response to the router receiving a message for an unknown destination host, and the message is passed through the network of routers until the destination host is located.

6. A digital communication system according to any previous claim wherein each router contains an LSP option memory for assembling, storing, and analyzing LSP options, the LSP option memory comprising a header section, an address section capable of storing a plurality of addresses, and a general section.

7. A method of operating a digital communication system comprising a network of routers linked together by links and having interfaces with local area networks (LANs) coupled to them, said method including the steps of:

operating the system under a protocol under which each LAN has a subnet address, and each host on a LAN has the subnet address as the high-order part of its own address;

providing each router with a set of interface/LAN tables listing the low-order address portions of the addresses of the hosts attached to the LAN plus the MAC (medium access control) identifiers of those hosts, and a set of link tables listing the subnet addresses of the LANs reachable through those links;

providing both the interface tables and the link tables in the routers with the full addresses of all hosts reachable through those interfaces and links;

each router, upon detection of ARP (address resolution protocol) requests from a source host for a destination host having the same subnet address as the source host but not on the same interface, returning a proxy ARP response giving the router's identification; and

providing each router with means for interrogating the interfaces for unknown hosts.

8. The method of claim 7 including the step of locating unknown hosts by a router by routine systematic polling of predetermined address space.

9. The method of claim 8 including the step of reserving a second predetermined address space for self-announcing hosts which address space is not systematically polled by a router.

10. The method of claim 7 including the step of initiating a poll for an unknown host in response to a router receiving a message for said unknown destination host, and passing the message through the network of routers until the destination host is located.

Patents Act 1977 Examiner's report to the Comptroller under Section 17 (The Search report)	Application number GB 9322910.2
Relevant Technical Fields (i) UK Cl (Ed.M) H4P (PPA, PPG) (ii) Int Cl (Ed.5) H04L 12/46, 12/66	Search Examiner MR J P COULES
	Date of completion of Search 7 DECEMBER 1994
Databases (see below) (i) UK Patent Office collections of GB, EP, WO and US patent specifications. (ii) ONLINE DATABASES: WPI, INSPEC	Documents considered relevant following a search in respect of Claims :- 1-10

Categories of documents

X:	Document indicating lack of novelty or of inventive step.	**P:**	Document published on or after the declared priority date but before the filing date of the present application.
Y:	Document indicating lack of inventive step if combined with one or more other documents of the same category.	**E:**	Patent document published on or after, but with priority date earlier than, the filing date of the present application.
A:	Document indicating technological background and/or state of the art.	**&:**	Member of the same patent family; corresponding document.

Category	Identity of document and relevant passages	Relevant to claim(s)
A, E	GB 2267418 A (ICL PERSONAL SYSTEMS OY) 1 December 1993 whole document	1 and 7

Databases: The UK Patent Office database comprises classified collections of GB, EP, WO and US patent specifications as outlined periodically in the Official Journal (Patents). The on-line databases considered for search are also listed periodically in the Official Journal (Patents).